Comment to: Forecast Rationality Tests Based on Multi-Horizon Bounds, by A. Patton and A. Timmermann

Barbara Rossi

Duke University

August 5, 2011

Abstract

This comment utilizes forecast rationality tests robust to unstable environments to revisit Patton and Timmermann's (2011) empirical evidence. The empirical results point to the presence of instabilities in the forecast rationality regression parameters.

Acknowledgements: I am grateful to A. Patton and J. Wright for sharing their codes and data, and to K. Hirano and J. Wright for organizing the special JBES session at the AEA.

Keywords: Forecasting, forecast optimality, real time data, forecast rationality tests, survey forecasts, forecast revisions, Greenbook

J.E.L. Codes: C22, C52, C53

Patton and Timmermann (2011) propose new and creative forecast rationality tests based on multi-horizons restrictions. The novelty is to consider the implications of forecast rationality jointly across the horizons. They focus on testing implications of forecast rationality such as the fact that the mean squared forecast error (MSFE) should be increasing with the forecast horizon (Diebold, 2001, and Patton and Timmermann, 2007) and that the mean squared forecast should be decreasing with the horizon. They also consider new regression tests of forecast rationality that utilize the complete set of forecasts across all horizons in a univariate regression, which they refer to as the "optimal revision regression" tests. One of the advantages of the proposed procedures is that they do not require researchers to observe the target variable, which sometimes is not clearly available. In fact, Patton and Timmermann (2011) show that both their inequality results as well as the "optimal revision regression" test hold when the short horizon forecast is used in place of the target variable. Their work is an excellent contribution to the literature.

The main objective of this comment is to check the robustness of forecast rationality tests to the presence of instabilities. The existence of instabilities in the relative forecasting performance of competing models is well-known (see Giacomini and Rossi (2010) and Rossi and Sekhposyan (2010), among others; Giacomini and Rossi (2011) provide a survey of the existing literature on forecasting in unstable environments). First, we show heuristic empirical evidence of time variation in the rolling estimates of the coefficients of forecast rationality regressions. We then use Fluctuation Rationality tests, proposed by Rossi and Sekhposyan (2011), to test for forecast rationality, while, at the same time, being robust to instabilities. We also consider a version of Patton and Timmermann's (2010) optimal revision regression test robust to instabilities, which we will refer to as the "Fluctuation Revision" test. Finally, we discuss the empirical evidence.

We focus on the same data as in Patton and Timmermann (2010), which include the Federal Reserve "Greenbook" forecasts of quarter-over-quarter rates of change in GDP and the GDP deflator. The data are from Faust and Wright (2009), starting in 1980 and ending in 2002.

First, we consider the typical "full-sample" Mincer and Zarnowitz (1969) forecast rationality test. Let the target value to be forecasted at time t using information up to time $t-h$ be y_t and let the forecast be denoted by $y_{t|t-h}$. The Mincer and Zarnowitz (1969) regression is as follows:

$$y_t = \alpha + \beta y_{t|t-h} + \varepsilon_{t,h}, \; t = 1, ..., P, \tag{1}$$

where P is the number of out-of-sample forecasts, h is the forecast horizon and $\varepsilon_{t,h}$ is the residual. If the forecasts are unbiased, the constant α should be statistically insignificantly different from zero; if the forecasts are optimal, the slope β should be statistically insignificantly different from unity. The null hypothesis of forecast rationality is that $\alpha = 0$ and $\beta = 1$, jointly. Table 1 reports the results. The table shows that forecast rationality is rejected at the 5% significance level for the GDP deflator inflation at most horizons, and it is rejected at horizons 1 to 3 for GDP growth.

Table 1. "Full-sample" Mincer-Zarnowitz's (1969) Forecast Rationality Tests

	GDP Deflator Inflation			Output Growth		
h	α	β	Joint	α	β	Joint
0	1.1600	5.2183	8.9164	3.9022	2.2612	4.2239
	(0.5599)	(0.0736)	(0.0116)	(0.1421)	(0.3228)	(0.1210)
1	0.5753	0.3029	4.5608	13.5417	17.8056	18.0331
	(0.7500)	(0.8594)	(0.1022)	(0.0011)	(0.0001)	(0.0001)
2	0.5275	0.5369	9.3500	8.3291	8.5530	9.0285
	(0.7681)	(0.7646)	(0.0093)	(0.0155)	(0.0139)	(0.0110)
3	0.0277	1.4636	8.4619	6.5810	4.0920	7.2051
	(0.9862)	(0.4810)	(0.0145)	(0.0372)	(0.1292)	(0.0273)
4	0.1992	3.6437	12.2100	1.6620	0.7996	3.0507
	(0.9052)	(0.1617)	(0.0022)	(0.4356)	(0.6705)	(0.2175)
5	2.2151	9.1047	14.8061	0.1852	0.0045	1.5906
	(0.3304)	(0.0105)	(0.0006)	(0.9116)	(0.9977)	(0.4515)

Note: Full sample Mincer and Zarnowitz's (1969) regression, eq. (1). P-values based on HAC robust estimates (with bandwidth equal to 3) for testing $\alpha = 0$ (column labeled "α"), $\beta = 1$ (column labeled "β") and both $\alpha = 0$ and $\beta = 1$ (column labeled "Joint") in parentheses.

However, the estimates of α and β may not be stable over time. The presence of instability is a serious concern, since it would imply that typical forecast rationality tests are invalid. See Rossi (2005) for an intuitive discussion of why full sample tests are invalid in the presence of instabilities. To provide informal evidence, Figure 1 reports estimates of α and β in rolling regressions, using a window of 60 out-of-sample forecast observations. The x-axis is the time of the latest forecast included in the rolling regression sample. Figure 1, Panel A shows that the estimates of α and β in regression (1) for the GDP deflator forecasts are quite unstable over time: α is closer to zero and β is closer to one in the late 1990's than in the mid-1990's. Similarly, Panel B shows that parameter estimates for GDP growth forecasts are also quite unstable over time. This evidence is only suggestive, though, since it ignores parameter estimation uncertainty.

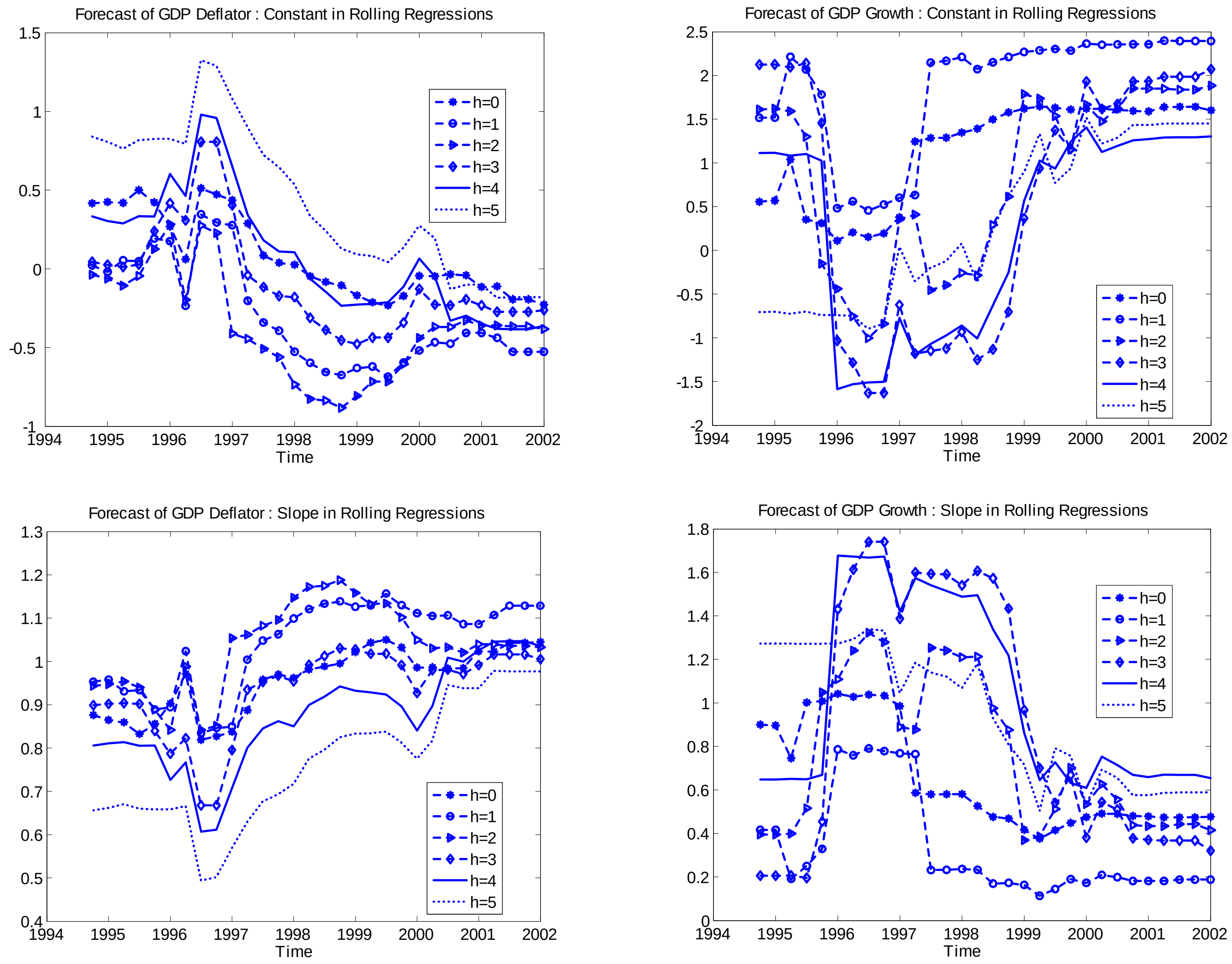


Figure 1, Panel A. GDP Deflator

Figure 1, Panel B. GDP Growth

Note. The figure shows rolling estimates of parameters of α and β in the Mincer and Zarnowitz (1969) regressions for various horizons (h).

In what follows, we will consider formal tests to investigate whether the empirical evidence in favor of the rejection of rationality in the Greenbook forecasts may depend on the sample period. We utilize the Fluctuation Rationality test developed by Rossi and Sekhposyan (2011), which is designed to test forecast rationality in unstable environments. Consider the general regression:

$$v_t = g_{t-h}' \cdot \theta + \eta_{t,h}, \ t = 1, ..., P, \tag{2}$$

where θ is a $(k \times 1)$ parameter vector, v_t is the realized variable, g_{t-h} is a $(p \times 1)$ vector of variables known at time $t - h$, and $\eta_{t,h}$ is the residual. Eq. (2) corresponds to (1) for

$\theta = [\alpha, \beta]'$, $v_t = y_t$, and $g_{t-h} = [1, y_{t|t-h}]$. Consider the following rolling regression. Let $\widehat{\theta}_t$ be the parameter estimate in regression (2) computed over centered rolling windows of size $m = 60$. That is, consider estimating regression (2) using data from $t - m/2$ up to $t + m/2 - 1$, for $t = m/2, ..., P - m/2 + 1$. Also, let the Wald test in the corresponding regressions be defined as:

$$\mathcal{W}_{t,m} = \left(\widehat{\theta}_t - \theta_0\right)' \widehat{V}_{\theta,t}^{-1} \left(\widehat{\theta}_t - \theta_0\right), \text{ for } t = m/2, ..., P - m/2 + 1, \tag{3}$$

where $\widehat{V}_{\theta,t}$ is a HAC estimator of the asymptotic variance of the parameter estimates in the rolling windows. Rossi and Sekhposyan (2011) define the Fluctuation Rationality test as:

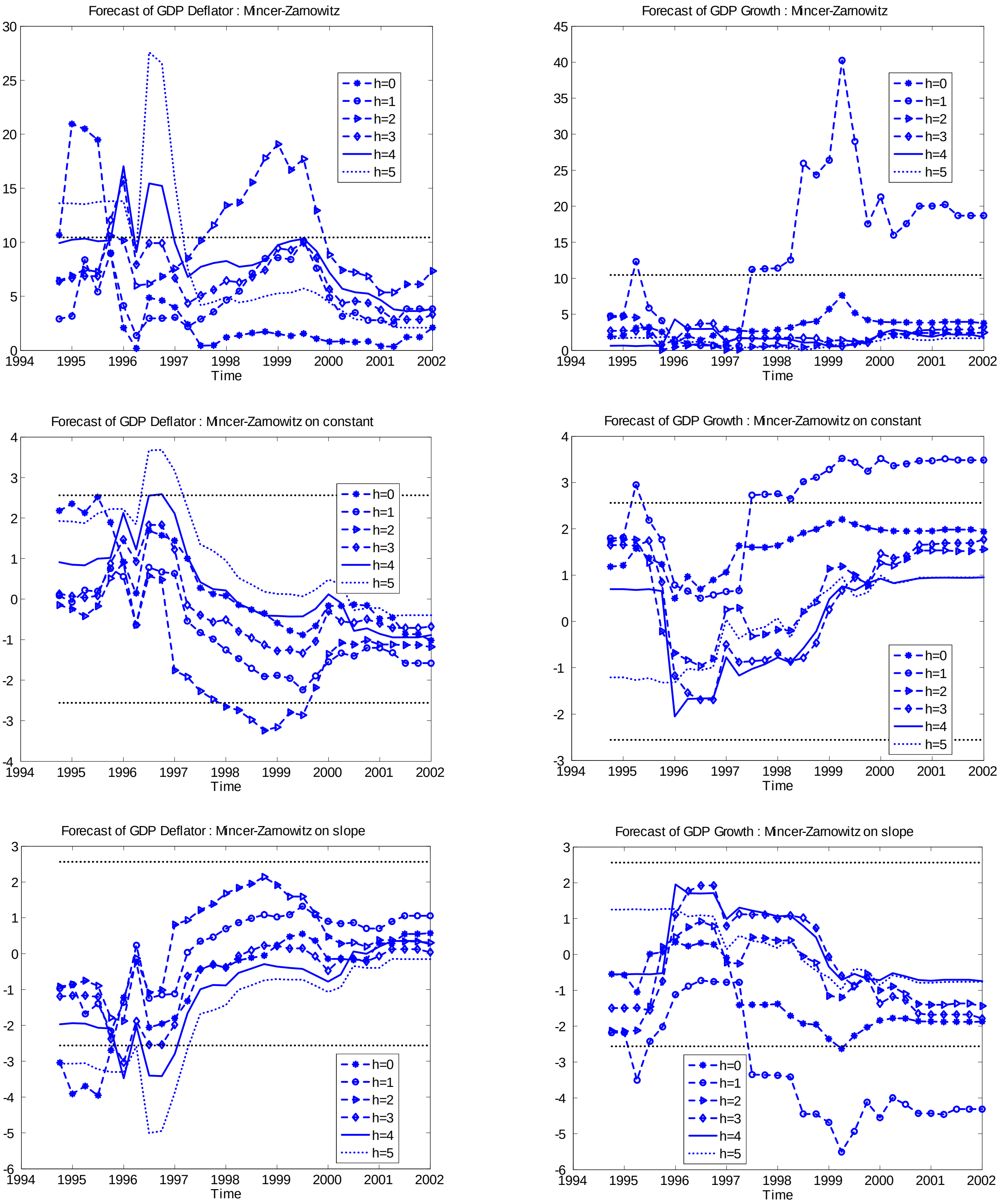
$$sup_t \mathcal{W}_{t,m}, \text{ for } t = m/2, ..., P - m/2 + 1. \tag{4}$$

The test rejects the null hypothesis $H_0 : E\left(\widehat{\theta}_t\right) = \theta_0$ for all $t = m/2, ..., P - m/2 + 1$ if $\max_t$ $\mathcal{W}_{t,m} > \kappa_{\alpha,k}$, where $\kappa_{\alpha,k}$ are the critical values at the $100\alpha\%$ significance level. The critical values at 5% are reported in Table 1 of Rossi and Sekhposyan (2011) for various values of $\mu = [m/P]$ and the number of restrictions, k.

A simple, two-sided t-ratio test on the s-th parameter, $\theta_0^{(s)}$, can be obtained as $\left(\widehat{\theta}_t^{(s)} - \theta_0^{(s)}\right)$ $\widehat{V}_{\theta^{(s)},t}^{-1/2}$, where $\widehat{V}_{\theta^{(s)},t}$ is element in the s-th row and s-th column of $\widehat{V}_{\theta,t}$. We reject the null hypothesis $H_0 : E\left(\widehat{\theta}_t^{(s)}\right) = \theta_0^{(s)}$ for all $t = m/2, ..., P - m/2 + 1$ at the $100\alpha\%$ significance level if $\max_t \left|\left(\widehat{\theta}_t^{(s)} - \theta_0^{(s)}\right) \widehat{V}_{\theta^{(s)},t}^{-1/2}\right| > \overline{\kappa}_\alpha$, where $\overline{\kappa}_\alpha$ are the critical values provided by Giacomini and Rossi (2010).

Figure 2 shows Fluctuation Rationality results for the Mincer and Zarnowitz (1969) regressions for the following cases: for testing $H_0 : \alpha = 0$ and $\beta = 1$ jointly for each horizon (first row); for testing $H_0 : \alpha = 0$ (second row); and for testing $H_0 : \beta = 1$ (third row). The figure also plots 95% confidence bands. The former is a one-sided test, whereas the latter are two-sided t-tests. More in details, the Mincer and Zarnowitz's regression is $W_{t,m} = \left(\widehat{\theta}_t - \theta_0\right)' \widehat{V}_{\theta,t}^{-1} \left(\widehat{\theta}_t - \theta_0\right)$ for $t = m/2, ..., P - m/2 + 1$ and $\theta_0 = [0, 1]'$ (first row); the Fluctuation rationality test on the constant is $W_{t,m}^{\alpha} = \widehat{\alpha}_t' \widehat{V}_{\alpha,t}^{-1/2}$ (second row) and that on the slope is $W_{t,m}^{\beta} = \left(\widehat{\beta}_t - 1\right) \widehat{V}_{\beta,t}^{-1/2}$ (third row), where $\widehat{V}_{\alpha,t}$ and $\widehat{V}_{\beta,t}$ are the diagonal elements of $\widehat{V}_{\theta,t}$. The figure shows that forecast rationality is rejected for horizons 0, 2 and 5 for the GDP Deflator, and for horizon 1 for the GDP growth.

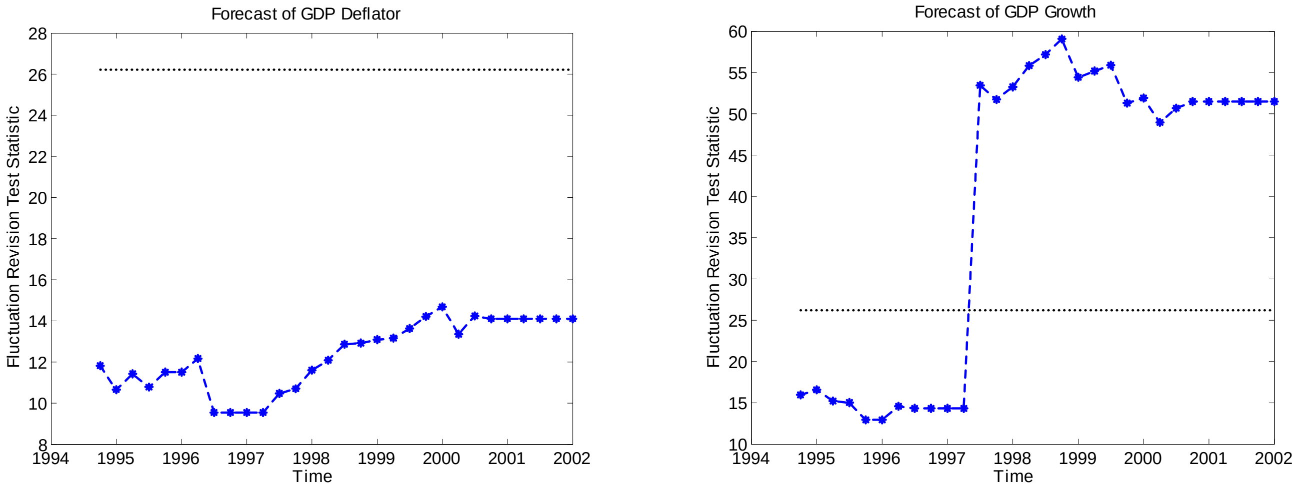
Figure 2. Panel A, GDP Defl. Panel B, GDP Growth



Note. The figure shows Fluctuation Rationality tests for the Mincer and Zarnowitz (1969) regressions for the following cases: for testing $H_0 : \alpha = 0$ and $\beta = 1$ jointly for each horizon (first row); for testing $H_0 : \alpha = 0$ (second row); and for testing $H_0 : \beta = 1$ (third row). In particular, the figure reports Equation (3) together with 95% confidence bands (dotted lines).

The framework discussed above can also be generalized to develop a version of the Patton and Timmermann's (2010) optimal revision regression test (implemented with a proxy) robust to instability. We refer to this test as the "Fluctuation Revision" test. The Fluctuation Revision test is defined as eq. (4), where $\mathcal{W}_{t,m}$ is defined by eq. (3), where $v_t = y_t$, $g_{t-h} = [1, y_{t|t-h_H}, d_{t|h_1,h_2}, ..., d_{t|h_{H-1},h_H}]$, H is the maximum forecast horizon, $d_{t|h_{H-1},h_H}$ denotes the forecast revision between horizons h_{H-1} and h_H, and $\theta_0 = [0, 1, ...1]'$. Figure 3 shows the test statistic, $\mathcal{W}_{t,m}$, over time; the dotted lines report the 5% critical value. According to the figures, the implications of forecast rationality considered by Patton and Timmermann (2010) are not rejected for the GDP Deflator, whereas they are rejected for GDP growth (mainly in the late 1990's).

Figure 3. Panel A, GDP Defl. Panel B, GDP Growth



Note. The figure shows the Fluctuation Revision test over time (line with stars) for GDP deflator data (left hand side panel) and GDP growth (right hand side panel), together with the critical value (dotted line).

To conclude, we found empirical evidence in favor of instabilities in the parameters of forecasting rationality regressions. Studying such instabilities might provide useful information as to when rejections of forecast rationality occurred, as well as their possible economic causes.

References

Diebold, F.X. (2001), *Elements of Forecasting*, South Western.

Faust, J. and J. Wright (2009), “Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset”, *Journal of Business and Economic Statistics* 27, 468-479.

Giacomini, R. and B. Rossi (2010), “Forecast Comparisons in Unstable Environments”, *Journal of Applied Econometrics* 25(4), 595-620.

Giacomini, R. and B. Rossi (2011), “Forecasting in Unstable Environments”, in: G. Elliott and A. Timmermann, *Handbook of Forecasting*, Vol. 2, Elsevier.

Mincer, J. and V. Zarnowitz (1969), “The Evaluation of Economic Forecasts,” in *Economic Forecasts and Expectations*, ed. J. Mincer, New York: National Bureau of Economic Research, 81–111.

Newey, W., and K. West (1987), “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, *Econometrica* 55, 703-708.

Patton, A. and A.Timmermann (2007), “Properties of Optimal Forecasts Under Asymmetric Loss and Nonlinearity”, *Journal of Econometrics* 140, 884-918.

Patton, A. and A.Timmermann (2011), “Forecast Rationality Tests Based on Multi-Horizon Bounds”, *Journal of Business and Economic Statistics*, forthcoming.

Rossi, B. (2005), “Optimal Tests for Nested Model Selection with Underlying Parameter Instabilities”, *Econometric Theory* 21(5), 962-990.

Rossi, B. and T. Sekhposyan (2010), “Have Models’ Forecasting Performance Changed Over Time, and When?", *International Journal of Forecasting* 26(4), 808-835.

Rossi, B. and T. Sekhposyan (2011), “Forecast Optimality Tests in the Presence of Instabilities", *mimeo, Duke University.*